In Our Shoes or the Protagonist's? Knowledge, Justification, and Projection[1]

Chad Gonnerman, Lee Poag, Logan Redden, Jacob Robbins, Stephen Crowley

Abstract

Sackris and Beebe (2014) report the results of a series of studies that seem to show that there are cases in which many people are willing to attribute knowledge to a protagonist even when her belief is unjustified. These results provide some reason to conclude that the folk concept of knowledge does not treat justification as necessary for its deployment. In this paper, we report a series of results that can be seen as supporting this conclusion by going some way towards ruling out an alternative account of Sackris and Beebe's results—the possibility that the knowledge attributions that they witnessed largely stem from *protagonist projection*, a phenomenon in language use and interpretation in which the speaker uses words that the relevant protagonist might use to describe her own situation and the listener interprets the speaker accordingly. With that said, we do caution the reader against drawing the conclusion too strongly, on the basis of results like those reported here and by Sackris and Beebe. There are alternative possibilities regarding what drives the observed knowledge attributions in cases of unjustified true belief that must be ruled out before, on the basis of such results, we can conclude with much confidence that the folk concept of knowledge does not treat justification as necessary for its deployment.

1. Introduction

That knowledge (K) entails justification (J) is widely accepted by philosophers. In view of the importance historically assigned to concepts in philosophy, we might ask, "Does the ordinary concept concur? Does the folk concept of knowledge treat justification as necessary for its deployment?" If it does, then, in cases where the protagonist's belief is clearly unjustified, we might expect to see a very strong *un*willingness on the part of ordinary people to attribute knowledge to the protagonist. Yet this is not what Sackris and Beebe (2014) observed. Across nine vignettes, they witnessed a sizeable tendency on the part of their participants to attribute knowledge in cases of unjustified true belief. On the basis of these results, we might conclude that the ordinary concept does not treat justification as necessary for knowledge, or, as we will often put it, that the K entails J thesis is not part of the folk concept.

But there are reasons to be wary of this inference. A common assumption in the cognitive sciences, one with which we operate in this paper, is that concepts are in-the-head structures that appear in a wide range of processes, including categorization processes (Machery, 2009; Murphy, 2004). It follows that, to be reasonably confident that the observed attribution rates show us that the folk concept of knowledge does not regard justification as necessary for its deployment, we must be reasonably sure that the attributions stem from the concept itself. Rival hypotheses must be ruled out. One of these is the possibility that the knowledge attributions

[1] This is a penultimate version of the paper to appear in *Oxford Studies in Experimental Philosophy* (Vol. 3). Please cite the final published version if possible. For helpful discussions, comments, and pointers, the authors thank Melissa Gonnerman, Luke Parker, audience members of the 2018 Buffalo Experimental Philosophy Conference, especially Josh Alexander, James Beebe, Kathryn Francis, Ángel Pinillos, and David Rose. We were also greatly aided by comments from the three anonymous reviewers and the editors of the volume.

observed by Sackris and Beebe stemmed largely from protagonist projection, a phenomenon in which a speaker uses words that the protagonist might use to describe her own situation and the listener interprets the speaker accordingly. Our results suggest that, however participants are engaging with Sackris and Beebe's materials, they are not interpreting the relevant knowledge claims projectively, at least not at appreciable rates. Thus, our results help to support Sackris and Beebe's conclusion to this extent: they go some way towards ruling out an alternative account of what drives the positive attributions that they witnesses. With that said, at the end of paper, we will stress various reasons for being wary about drawing the inference that the folk concept of knowledge does not treat justification as necessary for its deployment on the basis of results like those reported by Sackris and Beebe.

2. Background

Inspired by Sartwell's (e.g., 1991, 1992) attempts at dislodging the widespread assumption among philosophers that knowledge is more than true belief, Sackris and Beebe (2014) conducted a series of studies to explore whether the folk accept the K entails J thesis. Their results suggest an outsized willingness to attribute knowledge in two types of cases where justification is missing. The first are cases involving a true belief improperly grounded in a dream or a delusion. For instance, some participants received a vignette about Jordan who comes to believe that 125 is the square root of 15,625 because that is what the demonic voices in his head told him. Sackris and Beebe report that, across five cases of this sort, over 50% of their participants attributed knowledge. The second type of case explored by Sackris and Beebe is Sartwell's "knew it all along" cases. Here, participants received vignettes in which a protagonist forms a belief that gets vindicated, though at the time of formation the preponderance of evidence had by the protagonist speaks against the belief. For instance, in one vignette, participants learn about John who mulishly believes that his daughter is innocent of a crime in the face of strong incriminating evidence, with further evidence eventually coming in that corroborates his belief. In response to these vignettes, Sackris and Beebe report means around the scale's midpoint. This is a far cry from the basement-level responses we might expect if the folk concept of knowledge treats justification as necessary for its deployment.

While suggestive of a folk rejection of the K entails J thesis, work on the "cognitive aspects of survey methodology" gives some reason for pause. This work urges us to think of survey interactions as conversations, and thus as subject to principles of the sort that govern everyday interactions (e.g., Schwarz, 1996). Accordingly, a number of experiments indicate that participants sometimes interact with survey materials in ways that align with Grice's (1975) cooperative principle. One example comes from Schwarz, Strack, and Mai (1991). They report that correlations between participant assessments of how satisfied they are with their marriage and how satisfied they are with their life are weaker when participants first receive the marriage question than when they first see the life question. This finding can be interpreted in Gricean terms. The maxim of quantity prohibits conversational contributions that are either over- or under-informative. Lies of omission would be an example of the latter. If some such maxim guides participants when taking surveys, we might expect them to reinterpret the second of two questions if they see it as asking for information that is too similar to that given in response to the first. According to Schwarz et al., this is precisely what is happening in their study. The claim is that when participants see the marriage question followed by the life question, they tend

to re-interpret the latter in order to minimize repetition. In effect, the question becomes "Apart from your marriage, how satisfied are you with your life?" Other studies strike a similar theme. Yang (2005) identifies close to 30 studies of participants responding in ways that accord with Grice's maxims (see also Conrad, Schober, & Schwarz, 2014). It is thus reasonable to wonder to what extent conversational principles are at play in experimental philosophy. If they are, then the results of such studies might fail to reflect the significance assigned to them by experimental philosophers. We should expect as much in particular if conversational principles drive participants to interpretations of the relevant probes that go unrecognized by the experimenters.

To illustrate how Grice's maxims might get a grip in experimental philosophy, consider a cooperative participant assigned to one of Sackris and Beebe's delusion cases. After reading a story about an agent who believes everything the voices in his head tell him, including a random mathematical truth, the participant is asked, "Does the agent know this random claim?" If she is operating with something like Grice's maxim of quantity, she might balk at answering the question as literally interpreted. She might find herself (non-consciously) thinking, "It's just so obvious that the agent doesn't know, so it'd be uninformative for me to say as much." In a similar vein, when asked, "What did you do today?" unless joking, we don't say "Woke up!" because, as Schwarz (1999) notes while appealing to the maxim of quantity, that goes without saying. Maybe, in the minds of many, that the delusional agent does not know also goes without saying. If so, then our cooperative participant might go casting about for another interpretation of the question, one whose answer is less apparent. Or, to take a different route to the same possibility, suppose that ignorance is a norm of (some forms of) asking, as many claim (e.g., Hawthorne, 2004). Then, questions with perfectly clear answers should invite nonliteral readings. Such will happen, Whitcomb (2017) claims, if you ask someone who doesn't know that you are a philosopher whether the earth is the earth. According to him, regular folk are apt to think that you are joking or asking a question that calls for a metaphorical interpretation.

Assuming for the moment that conversational principles can lead to nonliteral interpretations of questions asked in experimental philosophy, what might a nonliteral interpretation look like in a study like Sackris and Beebe's? Holton (1997) reveals a possibility. Consider 'She sold him a pig in a bag. When he got home he discovered it was really a cat' (p. 627). Like the punch line in a joke that violates expectations created by the set-up, the second sentence forces a nonstandard interpretation of the first sentence. Given what the second sentence says, obviously, the swindler did not sell her buyer an actual pig. And, ultimately, we don't interpret it as if she did. Instead, on Holton's picture, what we do is we interpret the sentence as oriented to the buyer's point of view. It describes the transaction using words that he might have used to describe it at the time. A rough paraphrase of the interpretation we end up with might be "She sold him *what he thought at the time was* a pig in a bag." That this interpretation is a nonliteral interpretation of the original sentence is, we think, reasonably clear. As Jackson (2016, pp. 985-986) writes, commenting on Holton's 'Everyone was given a gold ring that turned out be brass', "I don't think the semantic theory should say that 'gold' sometimes means *gold*, and sometimes means *taken by some salient person x to be gold*. Rather, the latter meaning is sometimes recovered pragmatically as a conversational implicature." Much same seems true of the literal meaning of 'pig' and the interpretation that we end up with in response to the animal-in-the-bag sentence. The phenomenon at play here is what Holton calls *protagonist projection* (for detailed

discussion, see Stokke, 2013).[2] And he thinks that it can arise with uses of 'knows', as in 'She knew that he would never let her down, but, like all others, he did'. What we have here, according to Holton, is not a violation of factivity; it is a narrator putting himself in the shoes of the protagonist and reporting from her perspective.

Although embraced by many (e.g., Currie, 2010, p. 139n27; DeRose, 2009, pp. 16-17; Nagel, 2013, p. 286n11), not everyone has accepted Holton's account of non-factive uses of 'know' (e.g., Dahlman, 2017; Tsohatzidis, 1997). Still, that we sometimes adopt the perspective of others in narrative contexts has empirical support. Horton and Rapp (2003) report that response times to objects that are not visible from the perspective of a central character in a story are slower than to objects visible from her perspective. It is as if participants were "seeing" the objects through the character's eyes. Moreover, Buckwalter (2014) reports some telling results with respect to uses of 'knows'. He gave participants sentences of the sort that Hazlett (2010) uses to argue against the factivity of 'knows'. An example is 'Everyone knew that stress caused ulcers, before two Australian doctors in the early 80s proved that ulcers are actually caused by bacterial infection'. Buckwalter then asked participants to indicate which description best describes what was meant by 'everyone knew': (A) everyone thought they knew vs. (B) everyone really did know. 91% opted for (A), the projective reading.

The rival hypothesis that this empirical work presents for Sackris and Beebe's work is straightforward. When asked whether a belief grounded in delusion or dream, or a belief in a loved one despite the odds, amounts to knowledge, everyday conversational principles (or perhaps other mechanisms) may encourage participants to engage in protagonist projection. Many might end up interpreting the probe—the ascription to which participants indicate their agreement or disagreement—as capturing the very words that the protagonist might use to describe her situation. If so, then agreement to a probe so interpreted is not to assent to the proposition that the protagonist *knows*. It is to assent to something like the claim that the protagonist *thinks she knows*. Thus, if the assent is concept-driven—that is, primarily the product of conceptual processing operating over concepts retrieved from long-term memory, and possibly assembled "on the fly" (Barsalou, 1987)—then the assent won't stem from the participant's concept of knowledge. It will be her concept of thinking that one knows. And so, from this participant, we won't have evidence that her concept of knowledge rejects K entails J but that her concept of thinks that one knows rejects THK entails J.

3. Experiment 1

[2] To minimize confusion, there is a related phenomenon in the philosophical literatures from which we must distinguish protagonist projection. It is the curse of knowledge, also called epistemic egocentrism (Nagel 2010; for empirical studies, see Alexander, Gonnerman, & Waterman, 2014). The two are not the same. The curse of knowledge is a phenomenon that arises in mindreading. It involves projecting one's own appreciation of privileged information onto a relatively naive perspective, or, roughly speaking, mistakenly taking her to know what one knows. If Goldman (2006, pp. 165-166) is right, the curse of knowledge reflects a "quarantine violation" in the processes of constructing a model of the mental states of the mindreading target. Protagonist projection, on the other hand, is a phenomenon in language use and interpretation. In all likelihood, it stems from the pragmatics of language use and interpretation, "the study of meaning in context" (Scott-Phillips, 2015, p. xiii).

This study is an initial attempt to explore whether the rates of knowledge attributions observed in Sackris and Beebe's studies stemmed from protagonist projection to an appreciable degree.[3] The thought underlying this study is a simple one. If many participants did engage in protagonist projection in their studies, then, when given a response option that paraphrases the projective interpretation, those participants should prefer it to a simple knowledge attribution. So, after seeing vignettes of the sort used by Sackris and Beebe, we should see a decreased tendency to agree with a knowledge attribution of the form "S knows that *p*" when the other option is a projective paraphrase than when it is simply the option to disagree. Such, it seems, is the strategy for controlling for protagonist projection in experimental epistemology (e.g., Buckwalter, 2014; Machery, et al., 2017; Nagel, San Juan, & Mar, 2013; Rose, et al., forthcoming). In this study, we explore whether the prediction bears out.

3.1. Participants, materials, and procedure

Two hundred fifty-one participants were recruited through M-Turk. Prior to analysis, exclusion criteria were determined. Those who failed the attention check, who reported that they were not fluent in English, who did not complete the survey, or who may have taken the survey more than once as indicated by repeated IP addresses were cut from the main analyses of the paper. This determination resulted in a sample size of $N = 186$ (Age $M = 34.39$, Female = 41.9%).

All participants saw the following vignette from Sackris and Beebe:

> Brian is a 10-year-old boy who has just begun to study geometry. One night he goes to sleep and dreams that the square of the hypotenuse of a right triangle is equal to the sum of the squares of its other two sides. On the basis of this dream, he comes to believe the Pythagorean Theorem. A few days later in school his teacher introduces the Pythagorean Theorem for the first time in class. Brian thinks to himself "I already knew that the square of the hypotenuse of a right angle is equal to the sum of the squares of its other two sides."

After an attention question, participants in the first condition received a *standard probe*: "Please indicate whether you agree or disagree with the following claim: 'Brian already knew that the Pythagorean Theorem is true'. (i) Agree; (ii) Disagree." Participants in the second condition received a *projective probe*: "Which do you think best describes Brian in the story? (i) Brian thought he already knew that the Pythagorean Theorem is true; (ii) Brian really did already know that the Pythagorean Theorem is true". In each condition, the probe was followed by a question asking participants to indicate how confident or unconfident they were in their response using a fully anchored six-point scale ranging from "very unconfident" to "very confident".

3.2. Results and discussion

[3] It is unclear whether participants who agree with sentences of the form 'S knows that *p*' while assigning a projective interpretation to the sentences are best described as giving the experimenter a *knowledge* attribution. A more apt description might be that these participants gave something like a *thinks-one-knows* attribution. Perhaps, then, we shouldn't write about observed rates of knowledge attributions but rates of agreement to probes designed to assess whether participants think that the protagonist knows the relevant proposition. But would be quite cumbersome. Here, we will simply stick with 'knowledge attributions', using it to pick out indications of agreement to experimenter probes of the sort just described. As best as we can tell, nothing hinges on this stipulation.

A Pearson chi-square test revealed a statistically significant relationship between probe type and knowledge attributions: $\chi^2(1, N = 186) = 22.17$, $p < .001$, $\varphi = .35$). Participants were less likely to attribute knowledge when given the projective probe (40.4%) than when given the standard probe (74.7%) (see Figure 1). Binomial tests indicate that the proportion of participants attributing knowledge in the projective condition was marginally lower than .50, $p = .070$ while the proportion attributing knowledge in the standard condition was higher than .50, $p < .001$.

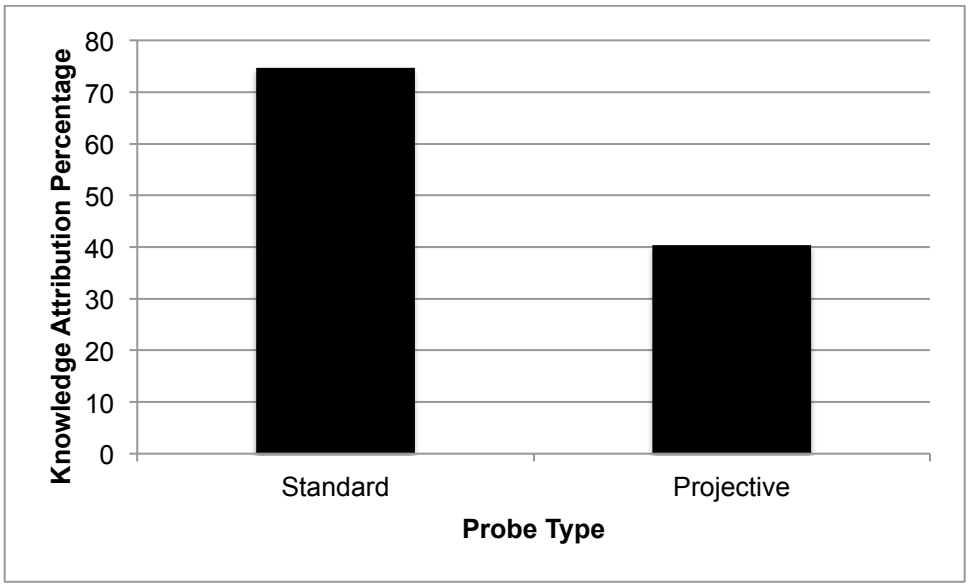

Figure 1. Percentage of knowledge attributions for standard probe and projective probe groups in Experiment 1

We also used a procedure from Starmans and Friedman (2012) to generate composite scores from responses to the knowledge probes and the confidence measures. Indications of agreement to the knowledge probes were coded as +1; disagreement, as -1. These scores were then multiplied by the corresponding confidence rating. For each participant, this gave us a *composite score*, ranging from -6 to +6. An independent samples *t*-test was then performed. It revealed a difference in composite scores across conditions (projective: $n = 99$, $M = -0.87$, $SD = 4.63$; standard: $n = 87$, $M = 2.36$, $SD = 4.07$): $t(184) = 5.06$, $p < .001$, $d = .75$. Moreover, a one sample *t*-test revealed that the composite score in the projection condition was marginally lower than the scale's midpoint while that in the standard condition was statistically higher than the midpoint (projective: $t(98) = 1.87$, $p = .065$; standard: $t(86) = 5.41$, $p < .001$).

This study revealed a decreased willingness to attribute knowledge when given a response option that paraphrases the projective interpretation. This is what we should find if many participants did engage in protagonist projection when given vignettes of the sort used by Sackris and Beebe. Thus, this study provides some preliminary evidence for thinking that the knowledge attributions observed by Sackris and Beebe stemmed partly from protagonist projection, though perhaps not completely, since over 40% of the participants attributed knowledge in the projective condition. It is reasonable to worry that this percentage is too high to support the projection hypothesis as articulated earlier—as the claim that the rates of attributions observed by Sackris and Beebe were

largely driven by protagonist projection. Still, it is only one study. Before any firm conclusions are drawn, we should see whether similar results emerge in other cases. The following study explores this by using Sartwell's "knew it all along" cases.

4. Experiment 2

4.1. Participants, materials, and procedure

Two hundred thirty-eight participants were recruited through M-Turk as part of another study. Exclusion criteria were similar to the previous study. The resulting sample size was $N = 173$ (Age $M = 35.78$, Female = 50.9%).

Participants were randomly assigned to one of two vignettes, which were taken from Sackris and Beebe.

> *Mickey*: The team of doctors responsible for treating Sandra's cancer told Sandra's husband, Mickey, that there was virtually no chance she should be able to beat the cancer and survive for more than a few months. In spite of what the doctors told him, Mickey was convinced that she would beat the cancer. In the end, Mickey's wife survived the cancer and remained cancer free for more than 35 years.
>
> *John*: John's daughter has been accused of murder. Even though she lacks a strong alibi and the police have compelling evidence against her, John feels that she must be innocent. After several very stressful weeks, the actual murderer finally comes forward and confesses.

After two attention questions, participants in the first condition received a *standard probe*. To illustrate, when it comes to *Mickey*, they were asked, "Do you agree or disagree that Mickey knew all along that his wife would survive the cancer?" Participants in the second condition received a *projective probe*. For example, "In your view, which of the following sentences better describes Mickey' situation? (i) Mickey really did know all along that his wife would survive the cancer; (ii) Mickey merely thought all along that he knew his wife would survive the cancer." These probes were then followed by a confidence measure of the sort used in the previous study.

4.2. Results and discussion

Table 1 shows the percentage of knowledge attributions in each of the four conditions. Binomial tests indicate that the proportion of participants attributing knowledge in the two projective conditions was lower than .50, $p < .001$ while the proportion attributing knowledge in the two standard conditions was higher than .50, $p < .001$.

Table 1
Percentages of Knowledge Attributions

	Standard	Projective	Row
Mickey	66.7	25.9	43.8
John	81.8	31.8	53.2

Column	73.3	28.6

A logistic regression was performed to explore the effects of probe type, story type, and their interaction on attributions of knowledge. The model was statistically significant, $\chi^2(3) = 37.94$, $p < .001$. It explained 26.3% (Nagelkerke R^2) of the variance and correctly classified 72.3% of cases.

As Table 2 shows, only the coefficient for probe type was significant. Neither a main effect of story type ($p = .521$) nor an interaction of probe type and story type ($p = .465$) was observed on knowledge attributions.

Table 2
Logistic Regression for Variables Predicting Knowledge Attributions

		95% C.I. for Odds Ratio		
	B(S.E.)	Lower	Odds Ratio	Upper
Probe (P)	2.27^{***} (0.56)	3.25	9.64	28.64
Story (S)	-0.29 (0.45)	0.31	0.75	1.81
P × S	-.52 (0.72)	0.15	0.59	2.41

Note. $^{***}p < .001$

Composite scores, calculated as before, reveal a similar pattern. We conducted a 2×2 ANOVA with probe type and story type as between-subject factors. The analysis revealed a main effect of probe type on composite scores, $F(1, 169) = 44.13$, $p < .001$, $\eta^2 = .20$. A main effect of story type was not observed, $F(1, 169) = 2.56$, $p = .112$, $\eta^2 = .01$, nor was an interaction between probe type and story type, $F(1, 169) = 0.13$, $p = .715$, $\eta^2 < .01$. In addition, one sample *t*-tests indicate that, in the standard conditions, the mean is statistically above the scale's midpoint ($t(74) = 4.43$, $p < .001$), while, in the projective conditions, the mean is statistically below the midpoint ($t(98) = 5.08$, $p < .001$).

Table 3 summarizes the mean composite scores and standard deviations.

Table 3
Mean Composite Scores and Standard Deviations

	Standard	Projective	Row
Mickey	1.57 (4.45)	-2.70 (4.45)	-0.83 (4.91)
John	2.91 (3.84)	-1.86 (4.71)	0.18 (4.94)
Column	2.16 (4.22)	-2.33 (4.56)	

It appears that a decreased tendency to attribute knowledge in projective conditions extends to "knew it all along" cases. This finding provides additional evidence that many participants in Sackris and Beebe's studies engaged in protagonist projection when responding to their knowledge probes.

5. Experiment 3

While the previous two studies generated results that were consistent with the projection hypothesis, and thus provided some evidence in favor of the hypothesis, certainly more evidence is preferable. There are other predictions worthy of exploration here. For instance, if participants are driven to projective interpretations because of Gricean mechanisms or an ignorance norm that they take to be operative in the survey context, we might expect a greater willingness to agree to the standard probe in cases of delusion when the belief is false than when it is true. After all, if the sheer obviousness of the lack of knowledge in the original cases, along with the thought that the questioner couldn't possibly be interested in something so obvious, is what drives participants to a protagonist projection, then we might expect that we will push more participants towards the projective interpretation if we make the lack of knowledge even more apparent. This should manifest itself as a *greater* inclination to agree with the standard probe when the belief is false than when it is true. This study looks into this possibility.

5.1. Participants, materials, and procedure

Ninety-one participants were recruited through Prolific. Exclusion criteria were the similar to the previous studies. This resulted in a sample size of $N = 75$ (Age $M = 31.23$, Female = 56.0%).

Participants were randomly assigned to one of two vignettes, which were based on a delusion case from Sackris and Beebe.

> *True Belief* [*False Belief*]: Ryan is a student, born and raised in Canada. He has recently become highly delusional. He claims that demons are talking to him inside his head and that they tell him all sorts of things. Ryan believes everything the demons tell him. One of the things the demons tell him is that Bill Morneau [Martin] is the Canadian Minister of Finance. Ryan has never followed Canadian politics very closely, but he comes to believe that Bill Morneau [Martin] is the Canadian Minister of Finance on this basis. It turns out, of course, that Bill Morneau really is [Martin is *not*] the current Canadian Minister of Finance.

Participants received a standard probe in both conditions: "Do you agree or disagree that Ryan knows that Bill Morneau [Martin] is the current Canadian Minister of Finance?" The probe was followed by an expanded but still fully anchored confidence measure, which now included a midpoint (1 = "very unconfident", 4 = "neither confident nor unconfident", 7 = "very confident"). Participants were then asked to respond to an attention check.

5.2. Results and discussion

A Pearson chi-square test found that participants were more likely to attribute knowledge in *True Belief* (72.2%) than in *False Belief* (20.5%): $\chi^2(1, N = 75) = 22.20$, $p < .001$, $\varphi = .52$ (see Figure 2).

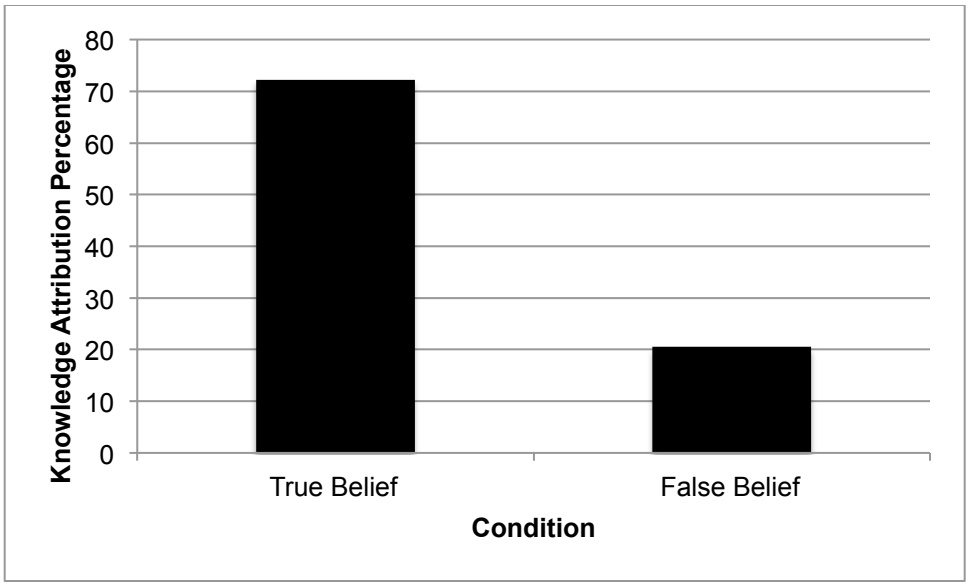

Figure 2. Percentage of knowledge attributions for true belief and false belief groups in Experiment 3

Looking at the composite scores, an independent samples *t*-test delivered a result similar to the chi-square test. They were larger in *True Belief* ($n = 36$, $M = 2.36$, $SD = 5.04$) than in *False Belief* ($n = 39$, $M = -3.67$, $SD = 4.49$): $t(73) = 5.48$, $p < .001$, $d = 1.28$.

These results appear to cut against the projection hypothesis, at least to the extent that it has been motivated by work on the cognitive aspects of survey methodology. If Gricean mechanisms or a norm of ignorance governing speech acts of asking drove many participants in Sackris and Beebe's studies to a projective interpretation of their knowledge probes, then, at first blush at least, it seems that we should have seen a greater tendency to attribute knowledge in *False Belief* than in *True Belief*. Again, if a literal response to the question of whether the protagonist knows in the case of an unjustified true belief is seen as setting up an answer that is seen as being too under-informative for a conversational contribution (and thus would amount to a violation the maxim of quantity) or seen as too obvious (and thus would be a violation a norm of ignorance on acts of asking), all the more should be true of an unjustified false belief. Yet this is not what we observed. In line of with other work in experimental epistemology (e.g., Nagel et al. 2013; Turri, Buckwalter, & Blouw, 2015), our results indicate that participants are less likely to attribute knowledge when an unjustified belief is false than when it is true.

Still, we shouldn't be hasty. Perhaps the results we observed can be explained in a manner consistent with the projection hypothesis. For instance, there may be something about the false belief case that suppresses, blocks, or bypasses the activation of Gricean machinery or an ignorance norm. At the moment, we are at a loss as to what this may be. But, given how little research has been done on the potential influences of conversational principles in experimental philosophy, our lack of ideas on this front should not carry much argumentative weight.

Another response to the results of this study is to hold on to the claim that projection is happening in Sackris and Beebe's studies but claim that the underlying mechanisms are neither

Gricean maxims nor a norm of ignorance governing certain forms of asking. Something else may be the driver. Although certainly possible, it is worth noting that adopting this response in the unfolding dialectic does come at a cost. It involves abandoning many of the considerations sketched in Section 2 for taking the projection hypothesis seriously in the first place. Absent these, the possibility that Sackris and Beebe's results stem from protagonist projection begins to look less probable. True, Experiments 1 and 2 can fill some of the evidential gap opened up by this dialectical move—but only some. Here, too, rival hypotheses concerning what the drives the observed effects must be ruled out. Perhaps the best dialectic move open to the person pushing the projection hypothesis is to find some nuance in the conversational principles at play in studies like Sackris and Beebe's that would predict the results just reported. We certainly welcome suggestions.

6. Experiment 4

The thought behind this study is that if a prominent tendency to engage in protagonist projection manifested itself in Sackris and Beebe's studies, then, when given cases of the sort that they used, along with suitable controls, an interaction effect should emerge. Specifically, we should see that there is a decreased tendency to attribute knowledge to the protagonist as we move from a standard to a projective probe that is greater in the case of an *unjustified* true belief than in a case a *justified* true belief. This study is an attempt to see whether this prediction bears out.

6.1. Participants, materials, and procedure

We used Prolific to recruit three hundred eighty-one participants. Exclusion criteria were the similar to the previous studies. This resulted in a sample size of $N = 288$ (Age $M = 32.13$, Female = 33.7%).

Participants were randomly assigned to one of four vignettes, which we based on two delusion cases from Sackris and Beebe.

> *Sunil Justified* [*Sunil Unjustified*]: Sunil is an exchange student who has recently become very studious [delusional]. He remarks to a friend that an esteemed professor of political science is teaching his class on American politics [that a malevolent demon from the netherworld is communicating with him inside his head]. Sunil believes most everything his professor tells him during class lectures [the demon tells him throughout the day]. One of the things his professor [the demon] tells him is that Mike Pompeo is the current U.S. Secretary of State. Sunil has never followed American politics very closely, but he comes to believe that Mike Pompeo is Secretary of State on this basis. It turns out, of course, that Mike Pompeo really is the current U.S. Secretary of State.
>
> *Jordan Justified* [*Jordan Unjustified*]: Jordan, a college-aged student, has recently become very studious [delusional]. He remarks to a friend that an esteemed professor of mathematics is teaching the class [that a demon is communicating with him inside his head]. Jordan believes most everything his professor tells him during class lectures [the demon tells him throughout the day]. One of the things that his professor [the demon] tells him is that 125 is the square root of 15,625, and he comes to believe that 125 is the

square root of 15,625 on this basis. It turns out that 125 really is the square root of 15,625.

Besides two attention checks, participants received either a standard probe or a projective probe. To illustrate, in the Sunil cases, the standard probe read: "Do you agree or disagree that Sunil knows that Mike Pompeo is the current U.S. Secretary of State?" The projective probe went: "In your view, which of the following sentences better describes Sunil's situation? (a) Sunil really does know that Mike Pompeo is the current U.S. Secretary of State; (b) Sunil only thinks he knows that Mike Pompeo is the current U.S. Secretary of State." The probes were then followed by a seven-point confidence measure of the sort used in Experiment 3.

6.2. Results and discussion

A logistic regression was performed to explore the effects of probe type, story type, and justification status on participants' attributions of knowledge. The model also included terms for the three two-way interactions. The model was statistically significant, $\chi^2(6) = 42.77$, $p < .001$. It explained 19.2% (Nagelkerke R^2) of the variance and correctly classified 70.1% of cases.

Table 4 shows that a main effect of probe type on knowledge attributions emerged, as did an interaction effect of probe type and story type. Neither a main effect of justification status ($p = .477$) nor of story type ($p = .863$) was observed. The same is true of an interaction of probe type and justification status ($p = .056$) and an interaction of justification status and story type ($p = .144$).

Table 4
Logistic Regression for Variables Predicting Knowledge Attributions

		95% C.I. for Odds Ratio		
	B(S.E.)	Lower	Odds Ratio	Upper
Probe (P)	3.08*** (0.70)	5.57	21.74	84.89
Justification (J)	0.30 (0.42)	0.59	1.35	3.05
Story (S)	-0.08 (0.45)	0.38	0.93	2.24
P × J	-1.20† (0.63)	0.09	0.30	1.03
P × S	-1.47* (0.66)	0.06	0.23	0.84
S × J	0.85 (0.58)	0.75	2.35	7.38

Note. †p <.10, *p < .05 ***p < .001

Composite scores were analyzed using a full-factorial 2×2×2 ANOVA with probe type, justification status, and story type as between-subject subject factors. The analysis revealed a main effect of probe type on composite scores, $F(1, 280) = 38.98$, $p < .001$, $\eta^2 = .12$. Neither a main effect of justification status nor a main effect of story type was observed (justification status: $F(1, 280) = 1.02$, $p = .315$, $\eta^2 < .01$; story type: $F(1, 280) = 0.05$, $p = .832$, $\eta^2 < .01$). An interaction effect of probe type and justification status did emerge, $F(1, 280) = 4.05$, $p = .045$, $\eta^2 = .01$. The remaining interactions were non-significant (probe type × story type: $F(1, 280) = 2.92$, $p = .088$, $\eta^2 = .01$; justification status × story type: $F(1, 280) = 1.53$, $p = .217$, $\eta^2 < .01$; probe type × justification status × story type: $F(1, 280) = 0.52$, $p = .471$, $\eta^2 < .01$).

Table 5 summarizes of the descriptives for the four conditions that remain after collapsing across the two types of stories.

Table 5
Mean Composite Scores and Standard Deviations Collapsed Across Story Types

	Standard	Projective	Row
Justified	3.58 (3.69)	-0.49 (4.85)	1.33 (4.80)
Unjustified	3.16 (3.60)	1.07 (4.65)	2.05 (4.31)
Column	3.35 (3.63)	0.34 (4.79)	

Analyses of simple effects revealed that participants were less inclined to attribute knowledge in response to the projective probe than the standard probe in the justified and unjustified cases, $F(1, 280) = 30.81$, $p < .001$, $\eta^2 = .10$ and $F(1, 280) = 10.02$, $p = .002$, $\eta^2 = .04$. In addition, these analyses did not find that, in response to the standard probe, participants were less inclined to attribute knowledge in the unjustified cases than the justified cases, $F(1, 280) = 0.46$, $p = .499$, $\eta^2 < .01$, though participants were *more* likely to agree to the projective probe in the unjustified cases compared to the justified ones, $F(1, 280) = 5.07$, $p = .025$, $\eta^2 = .02$.

At first blush, these results are encouraging for the projection hypothesis: we found an interaction effect of probe type and justification status. Closer inspection, however, reveals that the interaction effect did not go in the predicted direction (see Figure 3). Again, if the protagonist projection hypothesis is true, then the decreased tendency to attribute knowledge to the protagonist as we move from the standard to the projective probe should be greater in the unjustified cases than in the justified cases. In our study, we saw the reverse. For example, the difference between the mean composite scores in the unjustified true belief cases was 2.09. In the justified true belief cases, it was 4.07.

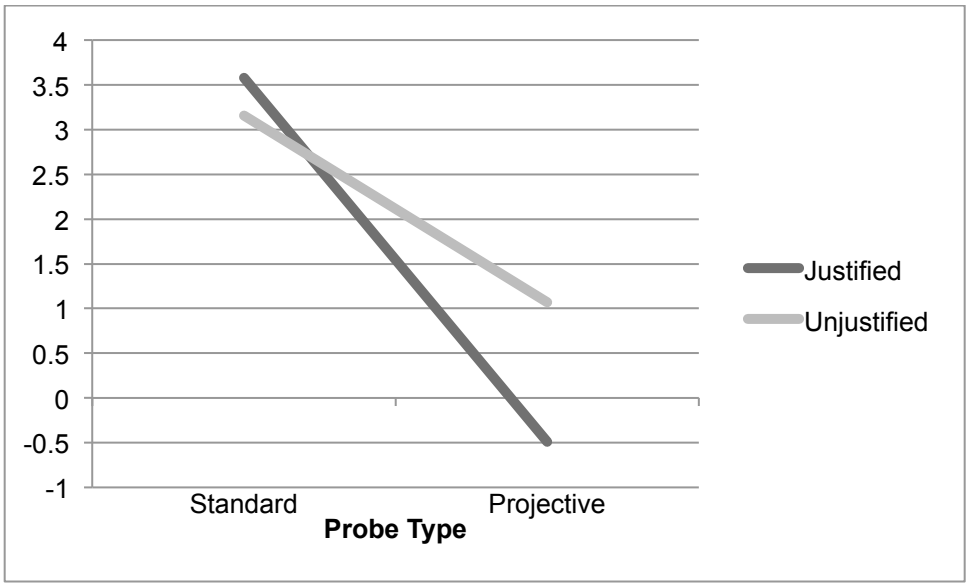

Figure 3. Probe type by justification status interaction on composite scores in Experiment 4

7. Conclusion

Sackris and Beebe report the results of a series of studies that puts some empirical pressure on the claim that the folk concept of knowledge treats justification as necessary for its deployment. It is reasonable, however, to wonder to what extent the knowledge attributions they observed largely stemmed from protagonist projection. On the whole, the results of this paper suggest that the extent is not all that extensive. Although the first two experiments reported here did deliver results that are consistent with the projection hypothesis, the last two did not. In our view, those two are more probative of the hypothesis, for two reasons. First, the theoretical and empirical frameworks that motivate Experiments 3 and 4 are largely independent of each other. So, the fact that both delivered disconfirmatory results amounts to a potent consideration against the hypothesis. Second, the results of Experiments 1 and 2, while in line with the projection hypothesis, are explicable in other ways. For instance, if something like the generalized context model (GCM) or its prototype equivalent—two formal models of categorization and category learning (for GCM, see Nosofsky, 2011; for the prototype model, see Minda & Smith, 2011)—extends to the folk concept of knowledge and the associated categorization processes, then it could be that the projective probe triggers a tightening of the concept's sensitivity parameter, thereby requiring a greater degree of similarity between the concept and the target for a positive verdict to emerge (Nosofsky & Johansen, 2000). Of course, only further empirical research will tell. Whatever the most plausible alternative explanation is, the projection hypothesis doesn't look so likely in view of the constellation of considerations just adduced. And so, when it comes to Sackris and Beebe's conclusion that the folk concept of knowledge does not treat justification as necessary for its deployment, our results help to strengthen their conclusion to this extent: they go some way towards ruling out an alternative account of the knowledge attributions that they observed.

With that said, there are various reasons to be wary about drawing the inference that folk concept of knowledge does not treat justification as necessary for its deployment on the basis of results like those reported by Sackris and Beebe. Many important questions remain unanswered. Experiments 1, 2, and 4 show that, in cases like those explored by Sackris and Beebe, how one asks the question about the protagonist's possible state of knowledge influences the rates of attributions observed. This result is similar to other results reported in experimental epistemology, including in Gettier, salience, and stakes cases (Buckwalter, 2014; Machery, et al. 2017; Nagel et al. 2013; Rose, forthcoming.). Which of the two ways of asking is better for probing the contents of the folk concept? If protagonist projection is occurring at appreciable rates, it is the projective probes because they help to control for unwanted interpretations of the relevant knowledge claims. If the GCM story just gestured at is right, the answer may be that both question formats are perfectly fine because the probe-driven dip in knowledge attributions simply reflects different similarity gradients used in applying the one concept to the vignette. And, surely, there are other possibilities as well. Ultimately, the answer to the question of which prompt format we should prefer when probing the folk concept of knowledge using vignettes partly depends on what explains the dip in knowledge attributions observed as we move from a standard to a projective probe. At the moment, we are largely in the dark as to what this explanation may be, at least when it comes to cases of the sort explored in this paper. But this is definitely an issue worthy of further exploration.

Moreover, before we draw any firm conclusions about the folk concept of knowledge from results like those reported here and by Sackris and Beebe, we must exert greater effort at controlling for or otherwise ruling out various forms of experimental noise. Experimentally determining whether the folk concept of knowledge treats justification as necessary for its application is in part the task of designing experiments that trigger responses that stem from categorization processes operating over this concept. If X% of our participants indicate that S knows that p in case C, we cannot simply assume that all of these indications have the appropriate etiology. To name just a few alternatives, some appreciable portion of the positive indications may reflect random or non-serious responding (Aust, Diedonhofen, Ullrich, & Musch, 2013), outcome bias (Gerken, forthcoming), background knowledge (Machery, 2009), yea-saying (Lavrakas, 2008), or demand characteristics (Powell, Horne, & Pinillos, 2014). Or, to cite one final complication, it may be that many of the attributions of knowledge observed here and by Sackris and Beebe come from participants who, to the chagrin of countless epistemologists, take the belief in question to be justified. The possibility is perhaps not as crazy as it might initially seem. After all, participants might find themselves thinking, what are the odds of, say, a delusional agent randomly picking out the one correct name out of many billions of possibilities for the current U.S. Secretary of State, unless he somehow picked up the name from his environment, maybe by walking past a newsstand or something? Naturally, attributions of knowledge in the absence of justification counts as evidence against K entails J being part of the attributor's concept of knowledge only if the attributor takes the case to be devoid of justification. Such are examples of the noise that we must control for or rule out before any firm conclusions are drawn about the contents of the folk concept on the basis of results like those reported here and in Sackris and Beebe's paper. So, yes, we agree that Sackris and Beebe's results, along with ours, do make the claim that the folk concept of knowledge rejects the K entails J thesis a decent bet at the moment. We just recommend exercising some caution when determining how much to place on the bet until further work comes out on matters like those discussed in this paragraph and the previous one.

Works Cited

Alexander, J., Gonnerman, C., & Waterman, J. (2014). Salience and epistemic egocentrism: An empirical study. In J. Beebe (Ed.), *Advances in experimental epistemology* (pp. 97-118). New York: Continuum.

Aust, F., Diedenhofen, B., Ullrich, & Musch (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods*, *45*, 527-535.

Barsalou, L. (1987). The instability of graded structure in concepts. In U. Neisser (ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 101-140). Cambridge: CUP.

Buckwalter, W. (2014). Factive verbs and protagonist projection. *Episteme*, *11*, 391-409.

Conrad, F. G., Schober, M. F., & Schwarz, N. (2014). Pragmatic processes in survey interviewing. In T. M. Holtgraves (Ed.), *The Oxford handbook of language and social psychology* (pp. 420-437). Oxford: OUP.

Currie, G. (2010). *Narratives and narrators: A philosophy of stories*. Oxford: Oxford University Press.

Dahlman, R. C. (2017). The protagonist projection hypothesis: Do we need it? *International Review of Pragmatics*, *9*, 134-153.

DeRose, K. (2009). *The case for contextualism*. Oxford: Oxford University Press.

Gerken, M. (forthcoming). Truth-sensitivity and folk epistemology. *Philosophy and Phenomenological Research*.

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford: OUP.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 41-58). New York: Academic Press.

Hawthorne, J. (2004). *Knowledge and lotteries*. Oxford: OUP.

Hazlett, A. (2010). The myth of factive verbs. *Philosophy and Phenomenological Research*, *80*, 497-522.

Holton, R. (1997). Some telling examples: A reply to Tsohatzidis. *Journal of Pragmatics*, *28*, 625-628.

Horton, W. S., & Rapp, D. N. (2003). Out of sight, out of mind: Occlusion and the acceptability of information in narrative comprehension. *Psychonomic Bulletin & Review*, *10*, 104-109.

Jackson, A. (2016). From relative truth to Finean non-factualism. *Synthese*, *193*, 971-989.

Lavrakas, P. J. (2008). Acquiescence response bias. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods* (pp. 3-4). Thousand Oaks, CA: SAGE.

Machery, E. (2009). *Doing without concepts*. Oxford: OUP.

Machery, E., Stich, S., Rose, D., Chatterjee, A., Karasawa, K., Struchiner, N., et al. (2017). Gettier across cultures. *Nous*, *51*, 645-664.

Minda, J. P., & Smith, J. D. (2011). Prototype models of categorization: Basic formulation, predictions, and limitations. In E. M. Pothos & A. J. Willis (Eds.), *Formal approaches in categorization* (pp. 65-87). Cambridge: CUP.

Nagel, J. (2010). Knowledge ascriptions and the psychological consequences of thinking about error. *Philosophical Quarterly*, *60*, 286-306.

Nagel, J. (2013). Knowledge as a mental state. In T. S. Gendler & J. Hawthorne (Eds.), *Oxford studies in epistemology* (Vol. 4, pp. 273-308). Oxford: Oxford University Press.

Nagel, J., San Juan, V., & Mar, R. A. (2013). Lay denial of knowledge for justified true beliefs. *Cognition*, *129*, 652-661.

Nosofsky, R. M. (2011). The generalized context model: An exemplar model of categorization. In E. M. Pothos & A. J. Willis (Eds.), *Formal approaches in categorization* (pp. 18-39). Cambridge: CUP.

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, *7*, 375-402.

Powell, D., Horne, Z., & Pinillos, N. Á. (2014). Semantic integration as a method for investigating concepts. In J. R. Beebe (Ed.), *Advances in experimental epistemology* (pp. 119-144). New York: Bloomsbury.

Rose, D., Machery, E., Stich, S., Alai, M., Angelucci, A., Berniūnas, R., et al. (forthcoming). Nothing at stake in knowledge. *Noûs*.

Sackris, D., & Beebe, J. R. (2014). In J. R. Beebe (Ed.), *Advances in experimental epistemology* (pp. 175-192). New York: Bloomsbury.

Sartwell, C. (1991). Knowledge is merely true belief. *American Philosophical Quarterly*, *28*, 157-165.

Sartwell, C. (1992). Why knowledge is merely true belief. *Journal of Philosophy*, *89*, 167-180.

Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic conversation*. Hillsdale, NJ: Erlbaum.

Schwarz, N. (1999). Self-reports: How questions shape the answers. *American Psychologist*, *54*, 93-105.

Scott-Phillips, T. (2015). *Speaking our minds: Why human communication is different, and how language evolved to make it special*. London: Palgrave Macmillan.

Starmans, C., & Friedman, O. (2012). The conception of knowledge. *Cognition*, *124*, 272-283.

Stokke, A. (2013). Protagonist projection. *Mind & Language*, *28*, 204-232.

Turri, J., Buckwalter, W., & Blouw, P. (2015). Knowledge and luck. *Psychonomic Bulletin and Review*, *22*, 378-390.

Tsohatzidis, S. L. (1997). More telling examples: A response to Holton. *Journal of Pragmatics*, *28*, 629-636.

Whitcomb, D. (2017). One kind of asking. *The Philosophical Quarterly*, *67*, 148-168.

Yan, T. (2005). *Gricean effects in self-administered surveys* (Unpublished doctoral dissertation). University of Maryland, College Park, MD.